CONTRIBUTIONS TO THE THEORY OF ACCIDENT PRONENESS

II. TRUE OR FALSE CONTAGION

BY

GRACE E. BATES and JERZY NEYMAN

CONTRIBUTIONS TO THE THEORY OF ACCIDENT PRONENESS

II. TRUE OR FALSE CONTAGION

BY

GRACE E. BATES AND JERZY NEYMAN

CONTRIBUTIONS TO THE THEORY OF ACCIDENT PRONENESS

II. TRUE OR FALSE CONTAGION

BY
GRACE E. BATES AND JERZY NEYMAN

1. Introduction. The first part of the present paper [1][1] was concerned with the theoretical aspect of the following practical question: can one use the number of light accidents incurred by different individuals in the past to predict the number of severe accidents in a hazardous occupation to be sustained in the future? The theoretical assumptions underlying this study form an extension of the well-known scheme due to Greenwood, Yule, and Newbold. An essential part of this scheme is characterized by the postulates: (*i*) that the individuals of a population differ from each other in accident proneness, (*ii*) that the accidents already incurred do not change the probabilities of further accidents in the future, and (*iii*) that these probabilities stay constant in time and are not modified by the experience that the individual may gain in the particular occupations. These three postulates may be symbolized by the combined term "mixture–no contagion–no time-effect model." In order to be able to deal with two kinds of accidents, light and severe, the above three postulates were supplemented by two more; (*iv*) that the expected number μ of light accidents per unit of time is proportional to the expected number λ of severe accidents (this postulate was termed the fundamental hypothesis), and (*v*) that to each severe accident there corresponds a fixed probability θ that the individual involved in the accident will survive.

The present Part II of the paper deals with a comparison between the foregoing scheme of mixture–no contagion–no time-effect and an alternative scheme due to Pólya [2]. Pólya's scheme postulates (I) identity of the individuals with respect to accident proneness—thus (I) is the denial of postulate (*i*)—, (II) possible presence of contagion of a specified type, and (III) possible effect of experience gained since entering the particular occupation. Curiously, as discussed by Lundberg [3] and Feller [4 and 5], if one considers the number of accidents incurred in a single period of time, its distribution implied by the mixture–no contagion–no time-effect model coincides with the distribution implied by the Pólya contagious scheme with the additional assumption that prior to the period of observation all the individuals concerned had the same number of accidents.

Naturally, no mathematical model of actual phenomena is ever absolutely exact. However, it is an undeniable fact that some models fit the particular set of phenomena better than some others. In the present case a number of problems concerned with personnel management make it important to distinguish between accidents which do and those which do not show elements of contagion in the sense of Pólya and between those in which the time elapsed since entering the hazardous occupa-

[1] Numbers in brackets refer to references at the end of the paper.

tion has some effect or no effect on the probability of accidents. This is just the problem studied in the following pages.

Using the scheme of Pólya in a slightly generalized form we deduce the multivariate distribution of the numbers of accidents of the same severe type incurred and survived in several successive periods of observation and of the number of these periods that are survived by the particular individual. This distribution is then compared with a corresponding distribution implied by the mixture–no contagion–no time-effect scheme. It is shown that, just as soon as the accidents are observed in more than one period, equivalence of the two distributions forces a condition on the parameters of the generalized Pólya scheme so that, barring an exceptional particular case, it is possible to distinguish between the two models.

As a by-product of this study we obtain the joint distributions of the number of light accidents incurred in the past and of the number of severe accidents to be incurred in the future, as implied by the postulates (*i*) through (*v*), which were given in Part I without proof. These distributions are deduced separately for those individuals who survive all the severe accidents incurred and separately for those who succumb.

In the last section we outline what seems to be a more promising method of approach to the problem of establishing the presence of contagion. This is based on the study of the distribution of time intervals between successive accidents incurred by particular individuals.

2. Basic assumptions. We shall consider an individual I who, from a certain moment $t = 0$ is exposed to the risk of accidents of a specified kind, subject to five postulates formulated below. The totality of these postulates will be denoted by (P) and described as the generalized Pólya contagious scheme or model. In formulating these postulates, it will be necessary to consider time intervals $(0, T_1)$ and (T_1, T_2) with $0 < T_1 < T_2$. These intervals will be always considered open on the left and closed on the right, say $0 < t \leqq T_1$ and $T_1 < t \leqq T_2$, where t stands for a moment in time.

POSTULATE P_1. *The individual I cannot die or otherwise cease to be exposed to accidents except as a result of an accident which may prove fatal.*

POSTULATE P_2. *Whatever the time interval (T_1, T_2) with $0 \leqq T_1 < T_2$, if the individual I is alive at T_1, the number of accidents, say $X(T_1, T_2)$, that he will incur and survive in (T_1, T_2) is a random variable whose distribution depends on T_1 and T_2 and on the number of accidents incurred in the time interval $(0, T_1)$, but not on the precise times when these accidents took place.*

Accordingly, we shall consider probabilities $P_{m,n}(T_1, T_2)$ and $Q_{m,n}(T_1, T_2)$ defined as follows:

$P_{m,n}(T_1, T_2)$ is the conditional probability that during the time interval (T_1, T_2) the individual I will incur exactly n accidents and that he will survive them all, given that at time T_1 he had incurred exactly m accidents and survived. If $T_1 = 0$, then the only acceptable value of m will be $m = 0$.

$Q_{m,n}(T_1, T_2)$ is the conditional probability that during the time interval (T_1, T_2) the individual I will incur exactly $n + 1$ accidents, that he will survive the first n and die in the $(n + 1)$st, given that at time T_1 he had incurred exactly m accidents and survived. Again, if $T_1 = 0$ then the only possible value of m will be $m = 0$.

Obviously,

$$(1) \qquad \sum_{n=0}^{\infty} \Big(P_{m,n}\,(T_1,\,T_2) + Q_{m,n}(\,T_1,\,T_2) \Big) \equiv 1 \; .$$

POSTULATE P_3. *If $T_2 \to T_1$ then all the probabilities $P_{m,n}(T_1,\,T_2)$ and $Q_{m,n}(T_1,\,T_2)$ converge to limits $P_{m,n}(T_1,\,T_1)$ and $Q_{m,n}(T_1,\,T_1)$, respectively.*

More specifically,

$$(2) \qquad P_{m,0}(T_1,\,T_1) = 1$$

for every m and, consequently, owing to (1)

$$(3) \qquad P_{m,n}(T_1,\,T_1) = Q_{m,n-1}(T_1,\,T_1) = 0$$

for $n \geqq 1$. The limits thus postulated will be interpreted as the probabilities of n accidents, all survived or not, occurring in the interval of time of zero duration.

POSTULATE P_4. *To each accident there corresponds a fixed probability θ of surviving it.* Consequently,

$$(4) \qquad Q_{m,0}(T_1,\,T_2) = (1 - \theta)\,[\,1 - P_{m,0}(T_1,\,T_2)]\;.$$

Remark: A superficial examination of the problem may suggest that instead of (4) the probability $Q_{m,0}(T_1,\,T_2)$ equals

$$(5) \qquad (1 - \theta)\,[\,P_{m,1}(T_1,\,T_2) + Q_{m,0}(T_1,\,T_2)]\;.$$

However, the reader will easily satisfy himself that the presumption is false because it does not take into account the circumstance that with the moment of a fatal accident the individual I ceases to be exposed to accidents which might have otherwise occurred after this accident.

POSTULATE P_5. *At least at $T_2 = T_1$, the probabilities $P_{m,n}(T_1,\,T_2)$ and $Q_{m,n}(T_1,\,T_2)$ are differentiable with respect to T_2 and, specifically,*

$$(6) \qquad \frac{\partial P_{m,\,0}\,(T_1,\,T_2)}{\partial T_2}\bigg|_{T_2 = T_1} = -\lambda\,\frac{1 + \mu m}{1 + \nu T_1}\;,$$

where λ, μ and ν are nonnegative constants, and

$$(7) \qquad \frac{\partial P_{m,\,n}\,(T_1,\,T_2)}{\partial T_2}\bigg|_{T_2 = T_1} = \frac{\partial Q_{m,\,n-1}\,(T_1,\,T_2)}{\partial T_2}\bigg|_{T_2 = T_1} = 0, \quad \text{for} \quad n \geqq 2\;.$$

It will be observed that with $\lambda > 0$, $\mu > 0$ and $\nu > 0$ equation (6) implies the contagion and the time effect. $P_{m,0}(T_1,\,T_2)$ represents the probability of avoiding accidents in $(T_1,\,T_2)$. When $T_2 = T_1$, then, according to (2), this probability is unity. Equation (6) implies that, with the increase of T_2 the speed of falling off in this probability is increased with the increase of m and is decreased with the increase

of T_1. Equations (7) imply that with $T_2 \to T_1$, the probability of more than one accident in (T_1, T_2) decreases faster than the difference $T_2 - T_1$. Also if $\mu = 0$, then there is no contagion. If $\nu = 0$, then there is no time effect.

The reader will notice that (4) and (6) imply

$$(8) \qquad \left. \frac{\partial Q_{m,\,0}\,(T_1,\,T_2)}{\partial T_2} \right|_{T_2=T_1} = (1 - \theta)\lambda \, \frac{1 + \mu m}{1 + \nu T_1}$$

and that then (1) and (7) imply

$$(9) \qquad \left. \frac{\partial P_{m,\,1}\,(T_1,\,T_2)}{\partial T_2} \right|_{T_2=T_1} = \theta\lambda \, \frac{1 + \mu m}{1 + \nu T_1} \,.$$

Pólya's original scheme was considered as a limiting case of a system of drawings from an urn and this led to the assumption $\mu = \nu$. Also, in the original scheme of Pólya $\theta = 1$, so that there is no room for the probabilities $Q_{m,n}(T_1, T_2)$.

As mentioned, the combination of postulates P_1 through P_5 will be denoted by (P) and described as the generalized Pólya contagious scheme. This scheme will be contrasted with another scheme to be denoted by (N) (connoting Newbold) which consists of postulates P_1 through P_5 supplemented by P_6 and P_7 as follows:

POSTULATE P_6. *Contagion and time effect are absent, so that $\mu = \nu = 0$.*

POSTULATE P_7. *The parameter λ in* (6), (8) *and* (9) *is a particular value of a random variable Λ with the probability density function*

$$(10) \qquad p_\Lambda(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \, \lambda^{\alpha-1} \, e^{-\beta\lambda} \qquad \text{for} \quad 0 < \lambda \,,$$

where α and β are arbitrary positive numbers.

It will be seen that except for the probability θ of surviving an accident, model (N) coincides with the original mixture–no contagion–no time-effect model of Greenwood, Yule, and Newbold.

Most of the study given below will refer to model (P) and it appears unnecessary to complicate the formulae with constant explicit references to this model. References to the two models in the form of letters P or N behind a vertical bar will appear only in cases when there may be a misunderstanding.

3. Problem studied. Considering model (P) we shall visualize s + 1 consecutive periods of time, the ith period beginning with $t_{i-1} \geqq 0$ and ending with $t_i > t_{i-1}$, where $t_0 = 0$ and $t_{s+1} = +\infty$. As before, these periods of time will be open on the left and closed on the right. With these periods of time we shall associate $s + 2$ random variables.

The random variable Z is defined as the number of complete periods of time survived by the individual I. Thus if $Z = 0$, then the individual I meets with a fatal accident in $(0, t_1)$, etc. If $Z = s + 1$ then the individual I survives up to $t_{s+1} = +\infty$. Obviously Z is capable of assuming integer values from zero to $s + 1$.

With each interval (t_{i-1}, t_i), where $i = 1, 2, \cdots, s + 1$, we associate a random variable X_i defined as the number of accidents incurred after the moment t_{i-1} and up to and including t_i, which the individual I will survive.

The variables X_i and Z are interdependent. Denote by k the value assumed by Z. If $i \leq k$ then X_i equals the number of accidents incurred by I in (t_{i-1}, t_i), all of which are survived. If $i = k + 1$, then X_i equals one less than the number of accidents incurred by I in (t_k, t_{k+1}), the last of these accidents being necessarily fatal. Finally, if $i > k + 1$, then $X_i = 0$.

Our problem is to deduce the joint probability generating function of the variables $Z, X_1, X_2, \cdots, X_{s+1}$.

Whatever be the random variables $Y_1, Y_2, \cdots, Y_r$ capable of assuming nonnegative integer values and whatever be the hypotheses H, we shall use the generic symbol

$$(11) \qquad G_{Y_1, Y_2, \ldots, Y_r} (u_1, u_2, \cdots, u_r \mid H)$$

$$= \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} \cdots \sum_{k_r=0}^{\infty} u_1^{k_1} u_2^{k_2} \cdots u_r^{k_r} \, P\{(Y_1 = k_1)(Y_2 = k_2) \cdots (Y_r = k_r) \mid H\}$$

$$= E \left(\prod_{i=1}^{r} u_i^{Y_i} \mid H \right)$$

to denote the conditional probability generating function of $Y_1, Y_2, \cdots, Y_r$, given the hypothesis H. Here the argument u_i corresponds to the variable Y_i and it is assumed that $|u_i| \leq 1$, for $i = 1, 2, \cdots, r$. For the variables $Z, X_1, \cdots, X_{s+1}$, the argument of the probability generating function which corresponds to Z will be denoted by v and the argument corresponding to X_i by u_i, $i = 1, 2, \cdots, s + 1$. With this notation, the object of our study is the generating function

$$(12) \qquad G_{Z, X_1, \ldots, X_{s+1}} (v, u_1, u_2, \cdots, u_{s+1} \mid P)$$

implied by model (P), its particular case

$$(13) \qquad G_{Z, X_1, X_2, \ldots, X_{s+1}} (v, u_1, u_2, \cdots, u_{s+1} \mid [\mu = \nu = 0] P)$$

and the counterpart of (12) implied by mixture–no contagion–no time-effect model (N). Obviously (13) is a function of λ and we have

$$(14) \qquad G_{Z, X_1, X_2, \ldots, X_{s+1}} (v, u_1, u_2, \cdots, u_{s+1} \mid N)$$

$$= \int_0^{\infty} G_{Z, X_1, \ldots, X_{s+1}} (v, u_1, \cdots, u_{s+1} \mid [\mu = \nu = 0] P) p_\Lambda (\lambda) \, d\lambda \ .$$

In the following we shall have occasion to use the fundamental relation between the absolute and the conditional expectation, familiar for a long time, but first rigorously established by Kolmogoroff [**6**]. Let $Y_1, Y_2, \cdots, Y_r$ be any random variables and let $f(y_1, y_2, \cdots, y_r)$ be any Borel measurable function of real arguments $y_1, y_2, \cdots, y_r$. Then

$$(15) \quad E[f(Y_1, Y_2, \cdots, Y_r)] = E\{E[f(Y_1, Y_2, \cdots, Y_r) \mid Y_1, Y_2, \cdots, Y_{r-1}]\} \ .$$

4. Preliminary formulae. Applying (15) we may write

$$(16) \qquad G_{Z,X_1,X_2,\ldots,X_{s+1}}(v, u_1, u_2, \cdots, u_{s+1}) = E\left[v^Z \prod_{i=1}^{s+1} u_i^{X_i} \right]$$

$$= E\left[v^Z E\left(\prod_{i=1}^{s+1} u_i^{X_i} \mid Z \right) \right]$$

$$= \sum_{m=0}^{s+1} v^m \, G_{X_1,X_2,\ldots,X_{s+1}}(u_1, u_2, \cdots, u_{s+1} \mid Z = m) P\{Z = m\}$$

Since $Z = m$ implies $X_i = 0$ for $i > m + 1$, the conditional probability generating function on the right of all the $s + 1$ variables $X_1, X_2, \cdots, X_{s+1}$ reduces to

$$(17) \qquad G_{X_1,X_2,\ldots,X_{m+1}}(u_1, u_2, \cdots, u_{m+1} \mid Z = m)$$

Our first step, then, will be to provide means for computing the probabilities $P\{Z = m\}$ and the conditional probability generating functions (17). For this purpose we return to the probabilities $P_{m,n}(T_1, T_2)$ and $Q_{m,n}(T_1, T_2)$ introduced in section 2. Multiplying them by u^n and summing for n from zero to infinity, we get, say

$$(18) \qquad g_m(T_1, T_2, u) = \sum_{n=0}^{\infty} u^n P_{m,n}(T_1, T_2)$$

and

$$(19) \qquad h_m(T_1, T_2, u) = \sum_{n=0}^{\infty} u^n Q_{m,n}(T_1, T_2) .$$

For $|u| \leqq 1$ both series converge and determine g_m and h_m as functions of u which are differentiable for $|u| < 1$. In many instances below the value of u will be immaterial and in these cases, to simplify the notation, we shall omit u from the symbol of the two functions. Also, whenever there is no danger of misunderstanding, we shall occasionally omit all three arguments and write simply g_m and h_m for the left-hand sides of (18) and (19).

In order to determine the functions g and h we proceed in the familiar manner [4 and 5] and write down the relation between $P_{m,n}(T_1, T_2)$ and $P_{m,n}(T_1, T_2 + \tau)$ where $\tau > 0$. We have

$$(20) \qquad P_{m,0}(T_1, T_2 + \tau) = P_{m,0}(T_1, T_2) P_{m,0}(T_2, T_2 + \tau)$$

and

$$(21) \qquad P_{m,n}(T_1, T_2 + \tau) = P_{m,n}(T_1, T_2) P_{m+n,0}(T_2, T_2 + \tau)$$
$$= P_{m,n-1}(T_1, T_2) P_{m+n-1,1}(T_2, T_2 + \tau) + o(\tau)$$

for $n > 0$, where, owing to (7), $o(\tau)$ decreases when $\tau \to 0$ and the rate of decrease is faster than that of τ. Subtracting $P_{m,0}(T_1, T_2)$ from both sides of (20) and $P_{m,n}(T_1, T_2)$

from both sides of (21), dividing the results by τ, passing to the limit as $\tau \to 0$, and using (6) and (9) we obtain

$$(22) \qquad \frac{\partial P_{m,0}(T_1, T_2)}{\partial T_2} = -\lambda \frac{1 + \mu m}{1 + \nu T_2} P_{m,0}(T_1, T_2)$$

and

$$(23) \qquad \frac{\partial P_{m,n}(T_1, T_2)}{\partial T_2} = -\lambda \frac{1 + \mu m + \mu n}{1 + \nu T_2} P_{m,n}(T_1, T_2)$$

$$+ \theta\lambda \frac{1 + \mu m + \mu(n - 1)}{1 + \nu T_2} P_{m,n-1}(T_1, T_2) .$$

Now we multiply (23) by u^n, sum for n from unity to $+\infty$, add to (22), use (18) and obtain, after some easy algebra,

$$(24) \qquad (1 + \nu T_2) \frac{\partial g_m}{\partial T_2} + \lambda\mu u(1 - \theta u) \frac{\partial g_m}{\partial u} = -\lambda(1 + \mu m)(1 - \theta u)g_m .$$

Using the familiar methods, the general solution of this partial differential equation is easily found to be

$$(25) \qquad u^{-(\gamma+m)} f\left[\frac{1 - \theta u}{u} A(T_2)\right]$$

where, to simplify the formula $\gamma = 1/\mu$ and, generally,

$$(26) \qquad A(T) = (1 + \nu T)^{(\lambda\mu)/\nu} .$$

Here $f(x)$ stands for an arbitrary differentiable function of the argument x. This function must be so selected that (25) coincide with $g_m(T_1, T_2)$. For this purpose we notice that the substitution $T_2 = T_1$ gives $g_m(T_1, T_1) \equiv 1$ identically in u and T_1. Making this substitution in (25) and equating the result to unity we obtain the condition determining the function $f(x)$,

$$(27) \qquad f\left[\frac{1 - \theta u}{u} A(T_1)\right] \equiv u^{\gamma+m} .$$

Now substitute

$$(28) \qquad \frac{1 - \theta u}{u} A(T_1) = x$$

and solve for u

$$(29) \qquad u = \frac{A(T_1)}{x + \theta A(T_1)} .$$

Substituting (29) into (27) we have

$$(30) \qquad f(x) = \left(\frac{A(T_1)}{x + \theta A(T_1)}\right)^{\gamma+m} .$$

Now the function $f(x)$ is determined. In order to obtain g_m we substitute (30) into (25). Easy algebra gives

$$(31) \quad g_m(T_1, T_2, u) = \left(\frac{A(T_1)}{\theta A(T_1) + (1 - \theta) A(T_2) + \theta[A(T_2) - A(T_1)](1 - u)} \right)^{\gamma + m}$$

$$= [D(T_1, T_2, u)]^{\gamma + m}, \quad \text{say.}$$

Now we turn to the function $h_m(T_1, T_2, u)$ generating the probabilities $Q_{m,n}(T_1, T_2)$. Using the same method we write

$$(32) \quad Q_{m,n}(T_1, T_2 + \tau) = Q_{m,n}(T_1, T_2) + P_{m,n}(T_1, T_2) Q_{m+n,0}(T_2, T_2 + \tau) + o(\tau)$$

and it follows

$$(33) \quad \frac{\partial Q_{m,n}(T_1, T_2)}{\partial T_2} = (1 - \theta)\lambda \frac{1 + \mu m + \mu n}{1 + \nu T_2} P_{m,n}(T_1, T_2) .$$

Multiplying this result by u^n, summing for n from zero to infinity and using (18) and (19) we obtain

$$(34) \quad \frac{\partial h_m}{\partial T_2} = (1 - \theta)\lambda \left[\frac{1 + \mu m}{1 + \nu T_2} g_m + \frac{\mu u}{1 + \nu T_2} \frac{\partial g_m}{\partial u} \right] .$$

The explicit expression of the derivative of h_m in (34) is obtained using (31). Since at $T_2 = T_1$, the value of h_m must be zero identically in u, an easy integration gives

$$(35) \quad h_m(T_1, T_2, u) = \frac{1 - \theta}{1 - \theta u} \left[1 - g_m(T_1, T_2) \right]$$

where g_m is given by (31).

Formulae (31) and (35) play a basic role in our further study. We begin by using (31) to evaluate the frequency function of the random variable Z.

5. Probability of surviving exactly j complete periods of observation. Referring to the definition of the function $g_m(T_1, T_2, u)$ and of the probabilities $P_{m,n}(T_1, T_2)$, it is easy to see that $g_m(T_1, T_2, 1)$ represents the conditional probability that the individual I will survive at least up to and including T_2, given that he was alive at T_1 and that up to the moment T_1 he sustained exactly m accidents. In particular, we obtain from (31)

$$(36) \quad g_0(0, T, 1) = [\theta + A(T)(1 - \theta)]^{-\gamma} = C_{0,T}^{-\gamma}, \text{ say,}$$

for the probability that the individual I, alive at $t = 0$ will survive at least up to and including an arbitrary moment $T \geqq 0$.

Now return to the random variable Z defined as the exact number of complete time intervals (t_{i-1}, t_i) which the individual I, alive at time zero, will survive. Whatever the nonnegative integer j, it is obvious that

$$(37) \quad P\{Z \geqq j\} = C_{0,t_j}^{-\gamma} = [\theta + A_j(1 - \theta)]^{-\gamma}$$

where, to simplify the notation, $A_j = A(t_j)$. It will be noticed that the conventional definition of $t_{s+1} = +\infty$ implies $A_{s+1} = +\infty$ and, therefore $P\{Z \geq s + 1\} = 0$. Now, the probability that I will survive exactly j complete intervals (t_{i-1}, t_i) is

$$(38) \qquad P\{Z = j\} = P\{Z \geq j\} - P\{Z \geq j + 1\}$$

$$= [\theta + A_j(1 - \theta)]^{-\gamma} - [\theta + A_{j+1}(1 - \theta)]^{-\gamma}$$

for $j = 0, 1, 2, \cdots, s$; while

$$(39) \qquad P\{Z = s + 1\} = 0 .$$

6. Probability generating function of Z, X_1, X_2, $\cdots$, X_{s+1} implied by model (P). In order to deduce the expression for the probability generating function desired we first establish a convenient recurrence formula.

Let $n_1, n_2, \cdots, n_j$ be any nonnegative integers. Define $S_0 = 0$ and generally

$$(40) \qquad S_i = \sum_{k=1}^{i} n_k .$$

Then the product

$$(41) \qquad \prod_{i=1}^{j} P_{S_{i-1}, n_i}(t_{i-1}, t_i)$$

represents the probability that the individual I, alive at time zero, will survive at least up to and including t_j, and that in the interval (t_{i-1}, t_i), with $i = 1, 2, 3, \cdots, j$, he will survive exactly n_i accidents. It follows that, by dividing (41) by $P\{Z \geq j\}$ we shall obtain the conditional probability of the compound event

$$(42) \qquad (X_1 = n_1)(X_2 = n_2) \cdots (X_j = n_j)$$

given that the individual I survives up to and including t_j. Thus

$$(43) \quad G_{X_1, X_2, \ldots, X_j}(u_1, u_2, \cdots, u_j \mid Z \geq j) = \frac{1}{P\{Z \geq j\}} \sum \prod_{i=1}^{j} u_i^{n_i} P_{S_{i-1}, n_i}(t_{i-1}, t_i)$$

where $\sum$ symbolizes j-fold summation for $n_1, n_2, \cdots, n_j$, each from zero to infinity. Referring to (18) and (31) we see that the last of these summations gives

$$(44) \quad \sum_{n_j=0}^{\infty} u_j^{n_j} P_{S_{j-1}, n_j}(t_{j-1}, t_j) = g_{S_{j-1}}(t_{j-1}, t_j, u_j)$$

$$= \left(\frac{A_{j-1}}{\theta A_{j-1} + (1-\theta)A_j + \theta(A_j - A_{j-1})(1 - u_j)} \right)^{\gamma + S_{j-1}}$$

$$= [D(t_{j-1}, t_j, u_j)]^{\gamma + S_{j-1}} .$$

We have then, in particular,

$$(45) \qquad G_{X_1}(u_1 \mid Z \geq 1) = \frac{g_0(0, t_1, u_1)}{P\{Z \geq 1\}} = \frac{g_0(0, t_1, u_1)}{g_0(0, t_1, 1)}$$

$$= \left[1 + \frac{\theta(A_1 - A_0)}{\theta + (1 - \theta)A_1}(1 - u_1)\right]^{-\gamma}.$$

Returning now to (43), if we multiply both sides of this equation by $P\{Z \geq j\}$ and use (44), we obtain

$$(46) \qquad P\{Z \geq j\}G_{X_1,X_2,\ldots,X_j}(u_1, u_2, \cdots, u_j \mid Z \geq j)$$

$$= [D(t_{j-1}, t_j, u_j)]^\gamma \left\{ \sum \prod_{i=1}^{j-1} (u_i D)^{n_i} P_{S_{i-1}, n_i}(t_{i-1}, t_i) \right\}$$

where, for short, $D = D(t_{j-1}, t_j, u_j)$ and the summation extends over all combinations of values of $n_1, n_2, \cdots, n_{j-1}$ from zero to infinity.

It is easily seen from (43) that the expression in curved brackets in (46) is equal to

$$(47) \qquad P\{Z \geq j - 1\}G_{X_1,X_2,\ldots,X_{j-1}}(u_1 D, u_2 D, \cdots, u_{j-1}D \mid Z \geq j - 1).$$

This establishes the recurrence formula sought, namely

$$(48) \quad G_{X_1,X_2,\ldots,X_j}(u_1, u_2, \cdots, u_j \mid Z \geq j)$$

$$= \frac{P\{Z \geq j-1\}}{P\{Z \geq j\}} [D(t_{j-1}, t_j, u_j)]^\gamma G_{X_1,X_2,\ldots,X_{j-1}}(u_1 D, u_2 D, \cdots, u_{j-1}D \mid Z \geq j-1).$$

Using this formula and (37), (45) we easily obtain

$$(49) \quad G_{X_1,X_2}(u_1, u_2 \mid Z \geq 2)$$

$$= \left\{1 + \frac{\theta}{\theta + (1 - \theta)A_2}[(A_1 - A_0)(1 - u_1) + (A_2 - A_1)(1 - u_2)]\right\}^{-\gamma}$$

and generally, by induction

$$(50) \qquad G_{X_1,X_2,\ldots,X_j}(u_1, u_2, \cdots, u_j \mid Z \geq j)$$

$$= \left[1 + \frac{\theta}{\theta + (1 - \theta)A_j} \sum_{i=1}^{j} (A_i - A_{i-1})(1 - u_i)\right]^{-\gamma}.$$

It is seen that the conditional distribution of $X_1, X_2, \cdots, X_j$, given that $Z \geqq j$ is always a j-variate negative binomial. We propose to call it the generalized Pólya multivariate distribution.

Now we can use the same method to compute the conditional distribution of $X_1, X_2, \cdots, X_{j+1}$ given that $Z = j$. To do so we turn our attention to (41) and notice that, if this product is multiplied by $Q_{s_j, n_{j+1}}(t_j, t_{j+1})$ then the result will equal the probability that the individual I, alive at zero time, will sustain and survive exactly n_i accidents in (t_{i-1}, t_i) for $i = 1, 2, \cdots, j + 1$ and that he will perish at the $(n_{j+1}+1)$st accident between t_j and t_{j+1}. It follows that

$$(51) \qquad P\{Z = j\} G_{X_1, X_2, \ldots, X_{j+1}}(u_1, u_2, \cdots, u_{j+1} \mid Z = j)$$

$$= \sum \prod_{i=1}^{j} u_i^{n_i} P_{S_{i-1}, n_i}(t_{i-1}, t_i) \sum_{n_{j+1}=0}^{\infty} u_{j+1}^{n_{j+1}} Q_{S_j, n_{j+1}}(t_j, t_{j+1}),$$

where the first sum extends over all values of $n_1, n_2, \cdots, n_j$ from zero to infinity.

However, referring to (19) and (35), we see that the last sum coincides with

$$(52) \qquad h_{S_j}(t_j, t_{j+1}, u_{j+1}) = \frac{1 - \theta}{1 - \theta u_{j+1}} [1 - g_{S_j}(t_j, t_{j+1}, u_{j+1})].$$

Substituting this result into (51), we have

$$(53) \qquad P\{Z = j\} G_{X_1, X_2, \ldots, X_{j+1}}(u_1, u_2, \cdots, u_{j+1} \mid Z = j)$$

$$= \frac{1 - \theta}{1 - \theta u_{j+1}} \left\{ \sum \prod_{i=1}^{j} u_i^{n_i} P_{S_{i-1}, n_i}(t_{i-1}, t_i) \right.$$

$$\left. - \sum \prod_{i=1}^{j} u_i^{n_i} P_{S_{i-1}, n_i}(t_{i-1}, t_i) g_{S_j}(t_j, t_{j+1}, u_{j+1}) \right\}.$$

Referring again to (43) and (44), we obtain easily

$$(54) \qquad P\{Z = j\} G_{X_1, X_2, \ldots, X_{j+1}}(u_1, u_2, \cdots, u_{j+1} \mid Z = j)$$

$$= \frac{1 - \theta}{1 - \theta u_{j+1}} \left\{ P\{Z \geqq j\} G_{X_1, X_2, \ldots, X_j}(u_1, u_2, \cdots, u_j \mid Z \geqq j) \right.$$

$$\left. - P\{Z \geqq j+1\} G_{X_1, \ldots, X_{j+1}}(u_1, \cdots, u_{j+1} \mid Z \geqq j+1) \right\}$$

which determines the generating function for the conditional distribution of $X_1, X_2, \cdots, X_{j+1}$ given that $Z = j$, for $j = 1, 2, \cdots, s$.

Substituting the explicit expressions (50) for the generating functions on the right of (54) and using (37) and (38) we obtain

$$(55) \quad P\{Z = j\} G_{X_1, X_2, \ldots, X_{j+1}} (u_1, u_2, \cdots, u_{j+1} \mid Z = j)$$

$$= \frac{1 - \theta}{1 - \theta u_{j+1}} \left\{ \left[\theta + (1-\theta)A_j + \theta \sum_{i=0}^{j} (A_i - A_{i-1})(1-u_i) \right]^{-\gamma} \right.$$

$$\left. - \left[\theta + (1-\theta)A_{j+1} + \theta \sum_{i=0}^{j+1} (A_i - A_{i-1})(1-u_i) \right]^{-\gamma} \right\}$$

where, with the convention $u_0 \equiv 1$ this formula is valid for $j = 0, 1, 2, \cdots, s$.

Finally, we have

$$(56) \quad G_{Z, X_1, X_2, \ldots, X_{s+1}} (v, u_1, u_2, \cdots, u_{s+1})$$

$$= \sum_{j=0}^{s} v^j \frac{1 - \theta}{1 - \theta u_{j+1}} \left\{ \left[\theta + (1 - \theta)A_j + \theta \sum_{i=0}^{j} (A_i - A_{i-1})(1 - u_i) \right]^{-\gamma} \right.$$

$$\left. - \left[\theta + (1 - \theta)A_{j+1} + \theta \sum_{i=0}^{j+1} (A_i - A_{i-1})(1 - u_i) \right]^{-\gamma} \right\}.$$

7. Probability generating function of $Z, X_1, X_2, \cdots, X_{s+1}$ implied by model (N). In the present section we use the results obtained to compute (14). For this purpose it will be sufficient to evaluate the limit

$$(57) \quad \lim_{\substack{\nu \to 0 \\ \mu \to 0}} P\{Z = j\} G_{X_1, X_2, \ldots, X_{j+1}} (u_1, u_2, \cdots, u_{j+1} \mid Z = j)$$

$$= F_j(\lambda, u_1, u_2, \cdots, u_{j+1}), \quad \text{say},$$

and to perform the integration

$$(58) \quad \int_0^\infty F_j(\lambda, u_1, u_2, \cdots, u_{j+1}) p_\Lambda(\lambda) d\lambda = F_j^*(u_1, u_2, \cdots, u_{j+1}), \quad \text{say}.$$

Then

$$(59) \quad G_{Z, X_1, X_2, \ldots, X_{s+1}} (v, u_1, u_2, \cdots, u_{s+1} \mid N) = \sum_{j=0}^{s} v^j F_j^*(u_1, u_2, \cdots, u_{j+1}).$$

To evaluate the limit in (57) we consider first the expression

$$(60) \quad \left[\theta + (1 - \theta)A_j + \theta \sum_{i=0}^{j} (A_i - A_{i-1})(1 - u_i) \right]^{-\gamma} = B_j, \quad \text{say}.$$

We have, recalling the definition of A_j in (26),

$$(61) \qquad \lim_{\nu \to 0} B_j = \left[\theta + (1 - \theta) e^{\lambda \mu t_j} + \theta \sum_{i=0}^{j} (e^{\lambda \mu t_i} - e^{\lambda \mu t_{i-1}}) (1 - u_i) \right]^{-(1/u)}$$

$$(62) \qquad \lim_{\substack{\nu \to 0 \\ \mu \to 0}} B_j = \exp \left\{ -\lambda \left[(1 - \theta) t_j + \theta \sum_{i=0}^{j} (t_i - t_{i-1}) (1 - u_i) \right] \right\}.$$

It follows, then, from (56) that

$$(63) \qquad F_j(\lambda, u_1, u_2, \cdots, u_{j+1})$$

$$= \frac{1 - \theta}{1 - \theta u_{j+1}} \left(\exp \left\{ -\lambda \left[(1 - \theta) t_j + \theta \sum_{i=0}^{j} (t_i - t_{i-1}) (1 - u_i) \right] \right\} \right.$$

$$\left. - \exp \left\{ -\lambda \left[(1 - \theta) t_{j+1} + \theta \sum_{i=0}^{j+1} (t_i - t_{i-1}) (1 - u_i) \right] \right\} \right).$$

Easy integration gives

$$(64) \quad F_j^*(u_1, u_2, \cdots, u_{j+1})$$

$$= \frac{1 - \theta}{1 - \theta u_{j+1}} \left\{ \left[1 + (1 - \theta) \beta^{-1} t_j + \beta^{-1} \theta \sum_{i=0}^{j} (t_i - t_{i-1}) (1 - u_i) \right]^{-a} \right.$$

$$\left. - \left[1 + (1 - \theta) \beta^{-1} t_{j+1} + \beta^{-1} \theta \sum_{i=0}^{j+1} (t_i - t_{i-1}) (1 - u_i) \right]^{-a} \right\}.$$

Substituting this expression for F_j^* in the right hand side of (59) we get the desired generating function.

It will be noticed from (59) that substituting into F_j^* unity for each of its arguments, we obtain the probability that $Z = j$ as implied by model (N). Thus

$$(65) \qquad P\{Z = j | N\} = [1 + (1 - \theta) \beta^{-1} t_j]^{-a} - [1 + (1 - \theta) \beta^{-1} t_{j+1}]^{-a}$$

and with the convention $t_{s+1} = +\infty$, it follows that

$$(66) \qquad P\{Z \geq j | N\} = [1 + (1 - \theta) \beta^{-1} t_j]^{-a}.$$

Dividing (64) by (65) we obtain a formula determining the conditional probability generating function of $X_1, X_2, \cdots, X_{j+1}$, given that in the interval (t_j, t_{j+1}) the individual I meets with a fatal accident,

(67) $P\{Z = j \mid N\}\ G_{X_1, X_2, \ldots, X_{j+1}}(u_1, u_2, \cdots, u_{j+1} \mid (Z = j), N)$

$$= \frac{1 - \theta}{1 - \theta u_{j+1}} \left\{ \left[1 + (1-\theta)\beta^{-1}t_j + \beta^{-1}\theta \sum_{i=0}^{j}(t_i - t_{i-1})(1 - u_i)\right]^{-a} \right.$$

$$\left. - \left[1 + (1 - \theta)\beta^{-1}t_{j+1} + \beta^{-1}\theta\sum_{i=0}^{j+1}(t_i - t_{i-1})(1 - u_i)\right]^{-a} \right\}$$

where again we adopt the convention $u_0 = 1$ so that (67) is valid for $j = 1, 2, \cdots, s$.

When comparing the models (N) and (P), formula (67) should be compared with (55). Both generate probabilities of the various combinations of values of the X_1, $X_2, \cdots, X_j$ and X_{j+1} subject to the restriction that in (t_j, t_{j+1}) the individual I meets with a fatal accident. In order to obtain for the model (N) the counterpart of formula (50) we notice that

(68) $\sum_{k=j}^{s} F_k^*(u_1, u_2, \cdots, u_{k+1}) \bigg|_{u_{j+1}=u_{j+2}=\ldots=u_{s+1}=1}$

$$= P\{Z \geqq j\} G_{X_1, X_2, \ldots, X_j}(u_1, u_2, \cdots, u_j \mid (Z \geqq j), N) .$$

Upon dividing by (66), we obtain

(69) $G_{X_1, X_2, \ldots, X_j}(u_1, u_2, \cdots, u_j \mid (Z \geqq j), N)$

$$= \left[1 + \frac{\beta^{-1}\theta}{1 + (1 - \theta)\beta^{-1}t_j} \sum_{i=1}^{j}(t_i - t_{i-1})(1 - u_i)\right]^{-a} .$$

8. Comparison between the distributions implied by models (P) and (N). The comparison between the implications of models (P) and (N) made thus far, [**3, 4,** and **5**], refer to the distributions of X_1 with $\theta = 1$. Using (50) and (69), we have

(70) $G_{X_1}(u_1 \mid (\theta = 1), P) = [1 + (A_1 - 1)(1 - u_1)]^{-\gamma}$

and

(71) $G_{X_1}(u_1 \mid (\theta = 1), N) = [1 + \beta^{-1}t_1(1 - u_1)]^{-a} .$

It is seen that, with $\alpha = \gamma$ and $\beta^{-1}t_1 = (A_1 - 1)$, the two distributions coincide so that no amount of empirical data regarding X_1 alone can afford means of distinguishing between the two models.

The above comparison is not entirely relevant, since it is frequently impracticable to ascertain the number of light accidents which the individuals of a population may have incurred prior to the period of observation. For this reason it is doubtful whether one could ever obtain data which could serve as an empirical counterpart

of the distribution generated by (71). The most one can hope to obtain is data regarding individuals who were exposed to unobserved accidents for approximately the same period of time, perhaps for a long time t_1, and then were subjected to observation during one or more subsequent periods (t_1, t_2), (t_2, t_3), etc. Such, for example, is true of the data on the London bus drivers discussed in Part I [**1**]. Before being employed by the London Transport Board, these 166 persons were experienced drivers and many of them must have had quite a few accidents which are not in the records. However, the time t_1 that elapsed between the obtaining of a driver's license and the beginning of the employment in London could probably be established with reasonable accuracy. Then, the statistics compiled for those drivers for whom t_1 has the same value could be used as a counterpart of the theoretical distributions of the random variables $X_2, X_3, \cdots$. We will compare these distributions for the two models (P) and (N), more generally, assuming that to each accident there corresponds a fixed probability θ of survival.

Consider first, then, formulas (50) and (69), with $u_1 = 1$. We have

$$(72) \qquad G_{X_2,\ldots,X_j}(u_2, \cdots, u_j | (Z \geq j), P) = \left[1 + \theta C_j \sum_{i=2}^{j} (A_i - A_{i-1})(1 - u_i) \right]^{-\gamma}$$

and

$$(73) \qquad G_{X_2,\ldots,X_j}(u_2, \cdots, u_j | (Z \geq j), N) = \left[1 + \beta^{-1}\theta C_j^* \sum_{i=2}^{j} (t_i - t_{i-1})(1 - u_i) \right]^{-a}$$

where, for simplicity

$$(74) \qquad C_j = [\theta + (1 - \theta)A_j]^{-1} \quad \text{and} \quad C_j^* = [1 + (1 - \theta)\beta^{-1}t_j]^{-1}.$$

It is seen that, if the observations are limited to one period only, e.g., from t_1 to t_2, then the distributions implied by the two models are single-variate negative binomials with two parameters each and are indistinguishable.

However, if the observations refer to two or more equal consecutive intervals, say $(t_i, t_{i+1} = t_i + 1)$ for $i = 1, 2, \cdots, j - 1$, then the situation is changed considerably. The coefficients of the binomials $(1 - u_i)$ in (73) are all equal to

$$(75) \qquad\qquad\qquad \beta^{-1} \theta C_j^*.$$

On the other hand, in (72) the coefficient of the binomial $1 - u_i$ is

$$(76) \quad \theta C_j (A_i - A_{i-1})$$

$$= \theta C_j \{[1 + \nu t_1 + \nu(i - 1)]^{(\lambda\mu)/\nu} - [1 + \nu t_1 + \nu(i - 2)]^{(\lambda\mu)/\nu}\}.$$

In order that the two distributions be forced to coincide by an appropriate choice of the parameters, it is necessary and sufficient that (76) be independent of i, that is to say, that

$$(77) \qquad\qquad\qquad \lambda\mu = \nu.$$

In other words, in comparing the joint distributions in models (P) and (N) of accidents survived in two or more equal consecutive periods up to and including, say, t_j, for those individuals who are known to be alive at t_j, we find that these joint distributions coincide if and only if condition (77) is met by the parameters λ, μ, ν of model (P). A comparison of formula (55) with (67), and (38) with (65) makes clear that the same conclusion is reached when one compares the joint distributions for those who are known to have succumbed to a fatal accident in the kth period, say, with $k = 1, 2, \cdots, j$, or when one compares the two distributions of the number of complete periods survived.

In principle, of course, this equality may be satisfied, and then the two schemes (P) and (N) will be indistinguishable no matter how many variables $X_2, X_3, \cdots, X_j$, we observe. However, the satisfaction of the equality (77) is most unlikely, and then the multivariate distribution implied by the model (P) will be different from that implied by model (N). At any rate, should the empirical distribution of $X_2, X_3, \cdots$, X_j indicate the inequality of the coefficients of the binomials $(1 - u_j)$ then this is an indication in favor of the model (P) rather than the model (N).

In the last section of the present paper we study the possibility of identifying the nature of contagion ("true" or "false") using a different set of observable random variables.

9. Joint distribution of the number of light and the number of severe accidents. In this section we use some of the formulae given above in order to deduce the joint distributions of the number Y of light accidents incurred in one period of time of unit length and the number X of severe accidents incurred in a subsequent period of time, also of unit length, as implied by the Greenwood–Yule–Newbold model supplemented by the fundamental hypothesis and by the assumption that to each severe accident there corresponds a fixed probability θ of surviving it. The formulae deduced here are given without proof in Part I of the present paper.

It will be realized that the process of determining the joint distribution of X and Y is exactly similar to that of section **7**.

The hypotheses assumed imply that for a given individual in the population, with a fixed proneness λ to severe accidents, the variables X and Y are mutually independent with the probability generating function of Y given by

$$(78) \qquad\qquad G_Y(v \mid \lambda) = e^{-A\lambda(1-v)}$$

where A is the modulus of frequency of the light accidents.

As to the variable X, we shall identify it with X_1 of the set of random variables discussed in section 7. Substituting u for u_1, i for t_i and unity for u_i with $i > 1$ in formula (63) we have

$$(79) \quad P\{Z = 0\}G_X(u \mid (Z = 0), N) = F_0(\lambda, u) = \frac{1 - \theta}{1 - \theta u}\left\{1 - e^{-\lambda(1-\theta u)}\right\}.$$

Similarly, we have from (63)

$$(80) \qquad P\{Z \geq 1\}G_X(u \mid (Z \geq 1), N)$$

$$= \sum_{j=1}^{s} F_j(\lambda, u, u_2, \cdots, u_{j+1})\Bigg|_{u_2 = \ldots = u_{j+1} = 1} = e^{-\lambda(1-\theta u)}.$$

Multiplying (79) by (78) we obtain, say

$$(81) \qquad \Phi_0(\lambda, u, v) = \frac{1 - \theta}{1 - \theta u} \left\{ e^{-A\lambda(1-v)} - e^{-\lambda[1-\theta u + A(1-v)]} \right\}$$

and similarly, multiplying (80) by (78),

$$(82) \qquad \Phi_1(\lambda, u, v) = e^{-\lambda[1-\theta u + A(1-v)]} .$$

Multiplying (81) and (82) by the probability density function of Λ and integrating for λ from zero to infinity we obtain, say

$$(83) \qquad \Phi_0^*(u, v) = \frac{1 - \theta}{1 - \theta u} \beta^a \{ [\beta + A(1-v)]^{-a} - [\beta + 1 - \theta + \theta(1-u) + A(1-v)]^{-a} \}$$

and

$$(84) \qquad \Phi_1^*(u, v) = \beta^a [\beta + 1 - \theta + \theta(1 - u) + A(1 - v)]^{-a}$$

respectively. Now, exactly as in section 7,

$$(85) \quad G_{X, Y}(u, v \mid (Z = 0), N) = \frac{\Phi_0^*(u, v)}{\Phi_0^*(1, 1)}$$

$$= \frac{1 - \theta}{1 - \theta u} \frac{[\beta + A(1 - v)]^{-a} - [\beta + 1 - \theta + \theta(1 - u) + A(1 - v)]^{-a}}{\beta^{-a} - (\beta + 1 - \theta)^{-a}}$$

and

$$(86) \quad G_{X, Y}(u, v \mid (Z \geqq 1), N) = \frac{\Phi_1^*(u, v)}{\Phi_1^*(1, 1)}$$

$$= \left[\frac{\beta + 1 - \theta}{(\beta + 1 - \theta) + \theta(1 - u) + A(1 - v)} \right]^a .$$

Formulae (85) and (86) are the two probability generating functions sought. The probabilities generated by (85) form the theoretical counterpart of the accident statistics for those individuals of the population who meet with a fatal accident. The probabilities generated by (86) correspond to the frequency distribution of accidents compiled for those who survive the second period of observation.

10. More hopeful approach to the problem of distinguishing between models (P) **and** (N). As was shown in section 8, the joint distribution of the numbers of accidents survived in several consecutive periods of observation implied by model (P) coincides with that implied by model (N) only in the improbable particular case when $\lambda\mu = \nu$. Thus, given a substantial number of observations of the simultaneous values of the random variables $X_1, X_2, \cdots, X_s$, it is possible to subject to a test the

hypothesis that the accidents considered are noncontagious and/or that there is no time effect. Details of such tests must be relegated to a separate paper. However, the authors anticipate that the power of the test contemplated will not be a very satisfactory one. On the other hand, it seems plausible that the power of a test of the same hypotheses may be much better if these hypotheses are tested on the observations, not of the variables $X_1, X_2, \cdots, X_s$ considered above, but of time intervals between successive accidents incurred by particular individuals. A complete study of this problem also requires more space than can be given in the present paper. However, it seems appropriate to include the present section outlining the new approach in relation to a small section of the statistical data that one may expect to have available, namely, in relation to those individuals who, during the period of observation, sustain exactly one accident. In this outline we shall assume $\theta = 1$. On the other hand, it appears possible to liberalize a little the original scheme of Pólya by not insisting that the dependence of the derivatives (6) on the number m of previous accidents is necessarily linear and by admitting that there may be a variation in accident proneness from one individual of the population to the next.

Consider an individual I and assume that from the moment $t = 0$ on, he is exposed to risk of accidents of a particular kind. For this individual we shall consider probabilities $P_{m,n}(T_1, T_2)$, defined in section 2, and shall assume that these probabilities depend on the number m of accidents sustained up to and including moment T_1 and also on the value of T_1 but not on the precise moments when the previous accident occurred. Specifically, we shall assume that

$$(87) \qquad \frac{\partial P_{m,n}(T_1, T_2)}{\partial T_2}\bigg|_{T_2 = T_1} = \begin{cases} -\dfrac{\lambda_m}{1 + \nu T_1}, & \text{if } n = 0 \\[2ex] \dfrac{\lambda_m}{1 + \nu T_1}, & \text{if } n = 1, \\[2ex] 0, & \text{if } n > 1 \end{cases}$$

where $\lambda_0, \lambda_1, \cdots, \lambda_m, \cdots$ are arbitrary nonnegative numbers and ν is subject to the restriction that for values of t limited to the period of observation $1 + \nu t > 0$.

Following the usual procedure, it is easy to find that

$$(88) \qquad P_{m,0}(T_1, T_2) = \left(\frac{1 + \nu T_1}{1 + \nu T_2}\right)^{\lambda_m/\nu}$$

and, using the assumption that $\lambda_m \neq \lambda_{m+1}$,

$$(89) \quad P_{m,1}(T_1, T_2) = \frac{\lambda_m}{\lambda_{m+1} - \lambda_m}\left[\left(\frac{1 + \nu T_1}{1 + \nu T_2}\right)^{\lambda_m/\nu} - \left(\frac{1 + \nu T_1}{1 + \nu T_2}\right)^{\lambda_{m+1}/\nu}\right].$$

Obviously, $P_{m,0}(T_1, T_2)$ is a decreasing function of λ_m. Thus, if all the λ's have the same value, the model implies the absence of contagion in the accidents. If the λ's increase,

$$(90) \qquad \lambda_0 < \lambda_1 < \cdots < \lambda_m < \lambda_{m+1} < \cdots,$$

then we shall speak of "regular positive" contagion, meaning that the more accidents the individual had in the past, the more intense is the risk of accidents in the future. If the λ's decrease,

$$(91) \qquad \lambda_0 > \lambda_1 > \cdots > \lambda_m > \lambda_{m+1} > \cdots,$$

then we shall speak of "regular negative" contagion. This would be the case where previous accidents "teach" the individual how to avoid accidents in the future.

Finally there is the possibility of the sequence of the λ's being nonmonotone. In this case we shall speak of "irregular contagion."

Now we shall assume that the individual I is observed for a unit of time, from T_1 to $T_1 + 1$. The object of this section is to deduce the conditional distribution of the random variable τ defined as the time between the beginning of the period of observation T_1 and the moment when the individual I sustains an accident, given that up to and including moment T_1 the individual sustained exactly m accidents and given that between T_1 and $T_1 + 1$ he sustains exactly one accident. Obviously $0 < \tau \le 1$.

For this purpose we compute the conditional probability, given exactly m accidents up to moment T_1, of the simultaneous occurrence of two events. One event is that during the period of observation the individual I sustains exactly one accident. The second event consists in τ not exceeding an arbitrary positive number $t \le 1$. The conditional probability just defined coincides with the conditional probability, given exactly m accidents up to moment T_1, that between T_1 and $T_1 + t$ the individual will have exactly one accident and that between $T_1 + t$ and $T_1 + 1$ he will have no accidents. Thus, this probability is easy to obtain from (88) and (89),

$$(92) \quad P_{m,1}(T_1, T_1 + t) P_{m+1,0}(T_1 + t, T_1 + 1)$$

$$= \frac{\lambda_m}{\lambda_{m+1} - \lambda_m} \left\{ \frac{a^{\lambda_m/\nu}}{(a+\nu)^{\lambda_{m+1}/\nu}} (a + \nu t)^{(\lambda_{m+1}-\lambda_m)/\nu} - \left(\frac{a}{a+\nu}\right)^{\lambda_{m+1}/\nu} \right\}.$$

Where, for the sake of simplicity

$$(93) \qquad a = 1 + T_1\nu.$$

Dividing (92) by the conditional probability of exactly one accident between T_1 and $T_1 + 1$, we obtain the conditional probability of $\tau \le t$, given that up to T_1 the individual I had exactly m accidents and that in $(T_1, T_1 + 1)$ he sustains exactly one of them. This probability is, then, the conditional distribution function of τ, say

$$(94) \qquad F_\tau(t \mid \psi_m, \nu) = \frac{P_{m,1}(T_1, T_1 + t) P_{m+1,0}(T_1 + t, T_1 + 1)}{P_{m,1}(T_1, T_1 + 1)}$$

$$= \frac{\left(1 + \frac{\nu t}{a}\right)^{\psi_m/\nu} - 1}{\left(1 + \frac{\nu}{a}\right)^{\psi_m/\nu} - 1}$$

where, for the sake of brevity, $\psi_m = \lambda_{m+1} - \lambda_m$. The corresponding probability density function is, say

$$(95) \qquad p_\tau(t \mid \psi_m, \nu) = \frac{\psi_m \left(1 + \dfrac{\nu t}{a}\right)^{\psi_m/\nu} - 1}{a\left[\left(1 + \dfrac{\nu}{a}\right)^{\psi_m/\nu} - 1\right]} \, .$$

We are particularly interested in the following three special cases obtainable from (95) by simple passages to the limit.

i) If $\psi_m = 0$, but ν is unspecified, then the mth accident is "noncontagious" and we have, say

$$(96) \qquad p_\tau[t \mid (\psi_m = 0), \nu] = \frac{\nu}{(a + \nu t) \log \left(1 + \dfrac{\nu}{a}\right)} \, .$$

ii) If $\nu = 0$ but ψ_m is unspecified, then there is no time effect but there may be contagion and

$$(97) \qquad p_\tau[t \mid \psi_m, (\nu = 0)] = \frac{\psi_m e^{\psi_m t}}{e^{\psi_m} - 1} \, .$$

iii) Finally, when both $\psi_m = 0$ and $\nu = 0$, then we have the no contagion–no time-effect case and

$$(98) \qquad p_\tau[t \mid (\psi_m = 0), (\nu = 0)] = 1 \, .$$

All five formulae (94) through (98) are given for $0 < t \leq 1$. They refer to a particular individual with a fixed number m of previous accidents and with fixed ψ_m, and it will be seen that their use is likely to give a substantial insight into the mechanism of accident proneness. The particularly interesting point is that, at least in some respects, the effect of variation from one individual of the population to another is now divided from contagion and time effect. Thus, if accident proneness conforms exactly with the no contagion–no time-effect mixture model of Greenwood, Yule, and Newbold, whether including the particular postulated distribution function of Λ or not, then the time intervals τ observed for arbitrary individuals of the population will be uniformly distributed between zero and unity as implied by (98). Any departure from this distribution is, then, an evidence of either time effect or contagion or both. Furthermore, the distribution (95) applicable in the general case coincides with (98) only when $\psi_m = \nu$.

The identity, with respect to any characteristic, of all individuals of a living population is always rather improbable. In particular it may be taken for granted that m will vary from one individual to another. Consequently, if by and large the accidents studied are subject to contagion and/or to time effect, it is most likely that at least for some individuals of the population the equality $\psi_m = \nu$ will not be satisfied and that, therefore, the study of the empirical distribution of τ will indicate the true nature of the machinery of accident proneness. This is particularly probable

in the two "regular" cases, in which the sequence of the lambdas is monotone. However, it may be hoped that a study of time intervals for individuals incurring two or more accidents during the period of observations will throw some light also on the irregular case in which ψ_m, owing to the variation in m, is positive for some individuals of the population and negative for others.

REFERENCES

[1] Grace E. Bates and Jerzy Neyman, "Contributions to the theory of accident proneness. I. An optimistic model of the correlation between light and severe accidents." *Univ. Calif. Publ. Statist.*, Vol. 1, pp. 215–254.

[2] G. Pólya, "Sur quelques points de la théorie des probabilités," *Ann. de l'Institut Henri Poincaré*, Vol. 1 (1930), pp. 117–161.

[3] Ove Lundberg, *On random processes and their application to sickness and accident insurance.* Uppsala, Almquist and Wiksells, 1940. 172 pp.

[4] W. Feller, "On the theory of stochastic processes, with particular reference to applications," *Proceedings, Berkeley Symposium on Mathematical Statistics and Probability.* Berkeley, University of California Press, 1949, pp. 403–432.

[5] W. Feller, *An introduction to probability theory and its applications.* New York, Wiley, 1950. xii + 419 pp.

[6] A. Kolmogoroff, *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Berlin, Julius Springer, 1933.